FROM: The Blackwell Guide to Research Methods in Bilingualism and Multilingualism
Editor(s):Li Wei & Melissa G. Moyer, 2008

10.1002/9781444301120.ch5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 Selecting Individuals, Groups, and Sites

Elizabeth Lanza

5.1 Introduction

This chapter focuses on the process of selecting the speakers or participants who will provide the data on linguistic structure and/or language use that are to be analyzed in a research project involving bilingualism. These individuals may belong to a group that is defined by, for example, language, e.g. Chinese in England (Li Wei, 1994), geography, e.g. Cape Town's District Six (McCormick, 2002), or linguistic disability. And within these larger groups, we may find other groups to which individuals belong by virtue of such variables as age, gender, socioeconomic status, language ideology, and so forth (see Heller, chapter 14 in this volume). Furthermore, the selection of the site for data collection from a particular group, for example, home or school, will have an impact on the study. Once the research question has been formulated (see Moyer, chapter 2 in this volume), the selection of the individuals, groups, and sites that will be the source of the relevant data will be decisive for the success of the study. Why is it so important to consider carefully the selection of individuals, groups, and sites? This chapter addresses the complexity of the issue.

The premise in this chapter is that bilingualism is socially embedded and that whether the focus of the study is on individuals, as in psycholinguistically oriented studies, or on groups, the study of bilingualism must take into account relevant sociolinguistic parameters. Cameron, Frazer, Harvey, Rampton, and Richardson (1992) go so far as to state that if we do research on language, we are social scientists. An important consideration in the social sciences, and more specifically the sociolinguistic study of language, is how the researcher affects the research process (cf. Milroy & Gordon, 2003). Data are not collected in a social vacuum. A focus on the bilingual's language use does not preclude interest in the actual languages involved, as in the investigation of the more formal aspects of code-switching. In this regard, the researcher's competence in these languages will play a part in the research design. And finally, many studies of bilingualism involve immigrants and although ethical considerations are relevant for all research, immigrant populations are particularly vulnerable.

The Blackwell Guide to Research Methods in Bilingualism and Multilingualism. Edited by Li Wei and Melissa G. Moyer
© 2008 Blackwell Publishing Ltd. ISBN: 978-1-405-12607-6

In what follows, I discuss the various considerations and implications involved in selecting individuals, groups, and sites for a study on bilingualism. I first highlight the importance of the selection process. Then I elaborate upon the issue of researcher identity and how that may bear directly on the research process, and particularly the selection process. Subsequently, I focus on the selection of individual participants in a research study focusing on individual bilingualism. Thereafter, I will discuss the selection of groups or communities of speakers as the point of departure for selecting individual participants. Even within a particular community or site, there is a need to consider variation, and the selection process has to take this into account. Finally, I bring up ethical considerations that are particularly relevant in bilingual research.

5.2 The Importance of the Selection Process

The selection of the participants and the site for a study is closely linked to the research design of the study (see Hua & David, chapter 6 in this volume). Every study will be motivated by a research question or hypothesis, the research design being the overall plan for addressing the issue in focus. The research design will have a theoretical motivation that includes a plan for the collection of the data to be used as the basis for investigating the research questions, for the operationalization of variables so that they can be measured, and for the analysis of the results. The research design will also provide information about who will comprise the study population, how they will be identified and contacted, and how informed consent will be sought (cf. Kumar, 1999). Paramount in this is that the procedures followed are adequate to obtain valid and reliable answers to the research questions. The research design is coherent in that theory, method, and data are all interconnected in a properly designed research study (see Moyer, chapter 2 in this volume).

The selection process is motivated by the study's objectives. For example, the starting point of a research question may be: What are the language choice patterns of non-Western second-generation immigrants in a Western society? Further specification of the research question will need to be made in order to help us in the selection process. Hence we may ask: What are the language choice patterns of children of Turkish immigrants in Denmark? But once again we will need to delimit the research question in order to restrict the selection process. This may be done by further specifying whether the Turkish immigrants have a Turkish or a Kurdish background, the age of the children, whether their language use patterns concern the home, school, or community club setting, whether gender is an issue, whether the children live in a larger metropolitan area or in a rural area, and so forth. A theoretical perspective that accentuates the importance of mapping variation will incorporate different types of speakers and different types of data used by the same speaker. In other words, the research question will assist us in planning the selection of individuals, groups, and sites from which to secure the necessary data.

10.1002/9781444301120.ch5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The extent to which we can generalize the findings from our research will depend to a significant degree on our selection process. If the individuals from whom we collect linguistic data are representative of larger groups to which they belong, we may be able to generalize our findings concerning the individuals to the larger group. This involves the issue of *representativeness* and a process called *sampling*. Sampling will be discussed in more detail in section 5.4. Suffice it to say here that the selection process is clearly related to, and a vital component of, the research design. And even when the focus is on individual bilingualism, the speakers will need to be selected according to carefully delineated profiles of bilingual competence, language use, and sociological variables.

The selection of participants for the study may affect other aspects of the research design, for example, the type of data that *can* be collected. If the population is an immigrant one involving older individuals who may not be literate, then for example an oral interview will be preferred to a written questionnaire, even though someone may be able to assist the individual in recording the responses (cf. Kroll, Gerfen, & Dussias, chapter 7 in this volume). This intervention could endanger the validity and reliability of the data. Awareness of these issues is important in our selection of research participants and data collection methods.

In the research question put forward above, Turkish immigrants in Denmark were proposed as examples of the broader category of non-Western immigrants in a Western society. However, one could have easily chosen Moroccan Arabic speakers in the Netherlands, or Filipinos in Norway. The particular selection process of a research study will invariably be related to the researcher's identity. This is the topic of the next section.

Section summary: Why is the selection of speakers and sites so important?

- It is an integral part of the research design of the study.
- It will affect the degree to which we can generalize our findings.
- It may affect other aspects of the research design, for example, the type of data that can be collected.

5.3 The Researcher's Identity in the Selection Process

The researcher's own identity, including his or her own particular interests, will greatly influence the research agenda. This identity is usually implicit in studies of bilingualism; however, as Li Wei (2000: 476) points out, we need to be aware of issues such as the researcher's linguistic competence, ethnicity, gender, age group, education level, disciplinary background, and attitude towards bilingualism. Indeed, one's gender identity may be decisive in, for example, a Muslim community where only women have access to Muslim women and their families. The researcher's

10.1002/9781444301120.ch5. Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

identity, as an insider or outsider to a community, may facilitate or complicate any attempts to overcome the "observer's paradox" (Labov, 1972c) – the problem that arises when the people being observed change their behavior in response to the observer's very presence. Some examples will illustrate how a researcher's identity can influence the research process, including the selection of participants and the site for research (see ideas on ethnography in Heller, chapter 14 in this volume).

Lanza (2004a) is a longitudinal study of the bilingual first language acquisition of two children in families in which the mother was American and the father Norwegian. As a bilingual speaker of English and Norwegian, I had the competence to analyze the language development of the children. Moreover, given the insight I had gained as the mother of bilingual children and my extensive network of bilingual families, the selection of this language and culture combination as the locus for my research was clear. I had a positive attitude towards bilingualism and I was personally motivated to learn more about how children acquire two languages simultaneously in a situation often referred to as "elitist" bilingualism (Harding-Esch & Riley, 2003). As an insider, I could observe how parents actually talked to their bilingual youngsters. A clear disadvantage of being an insider is that members of the group or community may not take seriously your questions and probes about the issue in focus for the research. Moreover, as I was undertaking research for my doctorate, my role as a researcher was hard to conceal despite my genuine interest in the mothers' attitudes and experiences. I was expected to have the answers for how to "successfully" bring up a child bilingually.

Zentella (1997), in her work on a New York Puerto Rican community, points out various dilemmas in being an insider to the community under investigation, especially when the group is a stigmatized one "which some readers wish to see vindicated and others hope to see chastised or rehabilitated" (p. 8). She states that, although there are definite advantages, "The closer the researcher is to the group, however, the more myopic the researcher may become about the significance of everyday acts that group members take for granted" (p. 7). Moreover, issues of power and solidarity may be at play. There are advantages and disadvantages to being an insider to a community. Hence identity as an insider is not decisive for the success of a study. Moyer (1998, 2000) investigated bilingual conversational data in Gibraltar, and was not an insider. However, not only was the researcher a bilingual speaker of English and Spanish, she was also familiar with language use patterns for communication in the community, as well as the social meanings associated with each language.

Some researchers are not insiders to the community selected and may in fact lack the necessary linguistic knowledge to engage in the project without assistance from within the group. Such studies are more often than not motivated by a need to gain more research-based knowledge about a particular community in order to provide better social or educational support for these individuals. Hence a need motivates the selection of a group, rather than the researcher's actual identity as an insider, or his or her competence in the minority language. Involving insiders in the research process can resolve the problem of the researcher's lack of linguistic competence; however, certain safeguards need to be taken. An example can illustrate this.

In Norway the largest group of minority language students come from Pakistan and have an Urdu/Panjabi background. Due to the lack of bilingual linguists with

10.1002/9781444301120.ch5. Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the necessary language competence, there has been an absence of studies of the language development and use of these children, who reportedly do poorly academically. The situation necessitates resorting to other methods in order to compensate for this lack, as demonstrated in Aarsæther (2004). He investigated negotiation skills among 10-year-old bilingual children of Pakistani origin in an Oslo school where the children used both Norwegian and Urdu/Panjabi effectively in their communication. Although the researcher was not competent in the minority language, he had had extensive experience in teaching children with that particular background, which enabled him to witness their language alternation in a school environment. In order to compensate for his lack of relevant linguistic skills, Aarsæther engaged the children's mother tongue instructor in the transcription and interpretation of the data. These interpretations were further cross-checked with other native speakers of Urdu/Panjabi to ensure validity and reliability. Hence the relationship in such a case between the researcher and the researched involved what Cameron, Frazer, Harvey, Rampton, and Richardson (1992) call "empowerment," in that individuals from the group under study are actually involved in the research process and in the interpretation of the results (cf. section 5.5).

The following example illustrates more explicitly how the selection process may be influenced by the researcher's identity and linguistic competence, all the while taking into account the theoretical motivations of the research design. Hvenekilde and Lanza (2001) is a study on language choice and social networks among Filipinos in Oslo. The actual choice of the city of Oslo was based on the fact that my colleague and I were residents of Oslo, and hence data collection would be easier to manage and accomplish. Moreover, we knew the city well. There was also a theoretical justification for this choice. Norway has become increasingly multilingual and multicultural despite images of Scandinavian homogeneity. And Oslo, with the greatest concentration of immigrants, fit the multilingual and multicultural profile sought after (see also Lanza & Svendsen, 2007).

The selection of the particular community of immigrants, the Filipinos, was not random. Many research manuals stress the importance of objectivity in research methods and that this requires a certain distance from the object of study. In recent years, however, there has been more social involvement in language-related issues among linguists as they attempt to address issues of social inequality related to language (cf. Wodak & Meyer, 2001; Fairclough, 2003a). What is important is that the researcher be aware of her or his ideological influences on the aims and objectives of the research. As Li Wei (2000: 479) states, ". . . bilingualism research can never be truly 'value-free.'" My colleague and I both had favorable attitudes towards bilingualism. I myself was an immigrant to Norway, having come as an adult, although as noted in Lanza (2004a: 74), the term *innvandrer* ('immigrant') has become reserved for those whose culture and language differ markedly from the Norwegian. Hence although I did not share the same ethnicity as members of the community in focus, I did share some common identity. My colleague and I were also both concerned about the negative media coverage that immigrants in Norway were receiving, focusing on the problems that certain immigrant groups had. Through our research we hoped to highlight some positive sides to immigration by focusing on a group that at the time had received little, if any, attention in the

10.1002/9781444301120.ch5. Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

media. Although several groups could have been selected, we chose the Filipinos. The selection of the Filipinos was also motivated by another aspect of my identity, as I had contact with this group through my participation in Oslo's Catholic cathedral parish where the Filipinos are the largest immigrant group.

The Filipinos also fulfilled the criteria that we had set up for selection. They maintained very close networks (Macdonald & Pesigan, 2000) and so provided an interesting case for testing theoretical claims concerning the impact of social networks on language maintenance (cf. Li Wei, 1994; see Xu, Wang, & Li Wei, chapter 15 in this volume). Our theoretical motivation was to investigate the relationship between ethnicity in social networks and language choice, particularly in the home with children, with an eye towards linguistic and cultural maintenance. Filipinos are multilingual before they arrive in Norway and thus have an extensive linguistic repertoire. Neither my colleague nor I had competence in any Filipino language; however, English is an official language in the Philippines and most Filipinos are competent in it. Moreover, these Filipino immigrants were competent in Norwegian to various degrees. To gain insider access to the community and to help us compensate for our lack of linguistic competence in Filipino languages, we employed the "friend-of-a-friend" fieldwork procedure (cf. Bossevain, 1974; Milroy & Gordon, 2003). My colleague knew a Filipino former student who had an extensive social network, and she became our field assistant. In section 5.4 below, I will return to this study and discuss how we actually selected the individual participants that we interviewed within the community, also addressing the issue of representativeness.

Section summary: The researcher's identity in the selection process

- The researcher's identity is an integral aspect of the research process, including the selection of individuals, groups, and sites.
- This identity includes the researcher's linguistic competence, ethnicity, gender, age group, education level, disciplinary background, and attitude towards bilingualism, as well as whether or not the researcher is an insider to the community.
- There are advantages and disadvantages to being an insider to the community you are studying.
- It is important that we are aware of our ideological influences on the aims and objectives of our research.

5.4 A Focus on Individual Bilingualism

Many studies in bilingual research focus on the bilingual individual and his or her competence in, or use of, two languages. In such cases, representativeness is not an issue as we are looking at human traits that are assumed to be universal. Laboratory

10.1002/9781444301120.ch5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

designs take the individual into account (see Kroll, Gerfen, & Dussias, and Abutalebi & Della Rosa, chapters 7 and 8 in this volume). The number of speakers is not as important as the criteria for selection. If we are focusing on variation in one particular aspect of bilingualism, and want to compare bilinguals, then we will need to match these bilinguals in respect of the other individual differences. As Grosjean (1998: 132) points out, although some research questions, for example those dealing with a bilingual's grammar, may be able to abstract away individual differences, most will not be able to do so. Hence these differences must be taken into account in our selection of speakers. Grosjean provides the most comprehensive list of basic individual differences that need to be accounted for when choosing bilingual speakers. These include the individual's language (acquisition) history and language use patterns, and the typological differences between the bilingual's languages. Other issues are how stable the languages are, and whether the bilingual is in the process of acquiring one or more languages, or losing them. To what extent is the bilingual proficient in each of his or her languages in each of the four skills: speaking, listening, writing, reading? Another issue concerns the bilingual's language modes. How often is the individual in a context that triggers a monolingual mode of language processing, or a bilingual mode? When in a bilingual mode, how often does code-switching take place? And finally, what are the traditional sociological variables represented by the bilingual, for example, age, gender, socioeconomic status, education, and so forth?

A crucial variable in the selection of bilingual children, which is also an important variable for the selection of adult participants in a study (cf. Hernandez, Bates, & Avila, 1994; Guion, Harada, & Clark, 2004), is the age at which the child acquired the two (or more) languages. The distinction is between bilingual first language acquisition and second language acquisition, otherwise referred to as simultaneous or infant bilingualism and consecutive or sequential bilingualism (cf. De Houwer, 1995). The study of each type of bilingualism poses different research questions. A traditional cutoff point between these two types has been three years, an age at which the child was considered essentially to have acquired his or her first language (McLaughlin, 1984). Bilingual first language acquisition research, however, has become increasingly stringent in its definition of the field: the child is expected to receive input from both languages from early on, indeed from birth (De Houwer, 1990). Input may come from interlocutors from outside of the family already from birth; nevertheless, the family has figured as the prominent source of input in available studies (cf. Lanza, 2004a, 2004b). Romaine (1995) set up a typology of six basic types of language use patterns in the home as a framework for selecting children for bilingual acquisition studies. These types varied according to the native language of the parents, the language(s) of the community, and the strategy the parents employ with the child. Some types render situations of simultaneous acquisition of two or even more languages (cf. Quay, 2001), others sequential acquisition. When selecting young bilingual children, the researcher needs to pay attention to the age at which the second language was introduced and to variation in the input.

Studies of young bilingual children are often case studies. Studies of bilingual aphasics and other brain-damaged individuals have also been case studies of individual patients, multiple case studies, or group studies (e.g. Gil & Goral, 2004, and

10.1002/9781444301120.ch5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the contributions in Paradis, 1995). In order to determine how many patients we will select for a study and what the selection criteria should be, we have to ask what the goal of the study is. For example, is the goal of the study to create a profile for patients with a particular injury, or to find appropriate measures for therapy? If the goal of the study is, say, to discover a linguistic structure typical for patients with Broca's aphasia, then a group study is necessary. A problem with a group study, however, is the lack of homogeneity among individual brain-damaged patients. In other words, in order to make a generalization about the group that is defined according to a particular syndrome, we must assume that the variation is trivial. Nonetheless, such variation has proven to be theoretically interesting. Furthermore, the border between the various syndromes may be vague, and the extreme cases are masked by any generalization. If the research question concerns a general phenomenon such as recovery patterns in bilingual aphasics (cf. Paradis, 2004: ch. 3), language proficiency and age of acquisition of each language can be important criteria for separating the participants. Individual and multiple case studies are common in bilingual language pathology; in fact, groups in such studies are often quite small. Individuals with brain damage are indeed an exceptional population and hence practical considerations will often be decisive in the selection of participants as we may not always be able to get exactly the type of speaker we wish for a study.

Whether dealing with bilingual brain-damaged patients or other bilingual speakers, our research design will determine whether or not we will need to select a control group of speakers. In a cause and effect research design we are measuring the impact that the independent variable has on the dependent variable (see Moyer, chapter 2 in this volume). However, we know that there are other extraneous variables that may be working in synergy with, or at odds with, the independent variable. In order to minimize the effect of these variables so that we may with greater confidence attribute the change in the dependent variable to the independent variable, we will need to select and set up a control group. For example, the study of the development of form and function in bilingual children has greatly profited from the use of the "frog story" (cf. Verhoeven & Strömqvist, 2001). The task itself, of letting the child browse through the pictures and then recount the story, may be a challenge to lexical retrieval and trigger hesitation. In order to measure the impact of these effects, we can establish a control group of monolingual children in each of the bilingual children's languages who will perform the same task. In this way the groups will be matched by the same social variables, with the marked variable separating them being that of bilingualism.

Section summary: A focus on individual bilingualism

- When focusing on the individual, the number of speakers is not as important as the criteria for selection.
- Some basic individual differences that must be accounted for when selecting speakers are the following:
  - speaker's language (acquisition) history,

10.1002/9781444301120.ch5. Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- language use patterns,
- the typological differences between the languages,
- whether the individual is in the process of language acquisition or attrition,
- the individual's proficiency in each of the four skills,
- the degree to which the individual is in a monolingual or bilingual mode of language processing,
- and finally, traditional sociological variables such as age, sex, socio-economic status, education, etc.

- Our research design will determine whether or not we need to select a control group of speakers in order to measure the effect of extraneous variables.

5.5 A Focus on the Group or Community

In a discussion on regional and immigrant minority languages, Extra and Gorter (2001) discuss the advantages and disadvantages of certain criteria that can be used to define and identify population groups in a multicultural society: nationality, birth country, self-categorization, and home language (see also Xu, Wang, & Li Wei, chapter 15 in this volume). In a study focusing on a particular bilingual population, once we have chosen the group or community for our research we are left with the selection of the individual participants from that group or community. Once the individuals are selected, the selection of the actual site for data collection is also important, as language used in different sites may represent society in different ways than correlations between individual and social variables such as age and gender.

In order to be able to say something about a community that is indeed *representative* for the community, we would ideally study the entire population. However, it is quite clear that such an endeavor would be a drain of human and financial resources. We need to resort to a sample, or subgroup, of the larger population that is our object of study. *Sampling* is the process of selecting a limited number of individuals from a group in order to estimate or predict aspects of the group. We may say that sampling is a compromise between accuracy and resources. If the sample is used as a basis for making statements about the entire community or population, the criteria for selecting the sample are paramount if we are to feel confident in the validity of our findings. Two factors may influence the degree of certainty about the inferences drawn from a sample: the size of the sample and the extent of variation in the sampling population – the greater the variation, the larger the sample needs to be. However, it is clear that practical considerations will also play a role in sample size. In order to avoid bias in the selection of the sample, we may resort to two types of sampling procedures: random/probability sampling designs and non-random/non-probability sampling designs.

A sampling design is a *random* or *probability sample* when each individual in the population has an equal and independent chance of inclusion in the sample (see Zhu & David, chapter 6 in this volume). If there are important social dimensions

10.1002/9781444301120.ch5. Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of variation such as age, gender, socioeconomic status, education, then a so-called *stratified* sample is needed which reflects these variables in the community. There are various methods in the social sciences for drawing a random sample, such as the fishbowl draw, computer programs, tables of random numbers, and picking, say, every tenth person. Ngom (2003) investigated lexical borrowing in the multilingual community on the island of St. Louis in Senegal where Arabic, French, and English each carry different prestige. All of the participants had Wolof as their native language. Two hundred participants were divided into two distinct age groups, each age group consisting of 50 males and 50 females. Ngom (p. 355) states, "Subjects were sampled randomly from houses on each of the major 20 streets and meeting places (where people from all over the island meet to spend time together) . . ." Thus age and gender were variables used to stratify the random sample. Why these variables were chosen was the researcher's judgment that they were relevant variables for the community under investigation (see below).

Large-scale variation studies may resort to random sampling. However, such methods assume that the universe from which the sample is drawn is known. There may be difficulty in achieving a random sample when dealing with bilingual minority communities as no relevant sampling frames or lists may exist that identify all of the individuals in the population. It would not be valid to use a telephone catalogue and pick out all persons that have, say, a Russian or a Chinese family name. Census information may try to document the number of people with a certain immigrant background but this is not always reliable. In immigrant communities, some members may be clandestine and wish to avoid public acknowledgment. Trying to map out bilingualism in a geographical area such as a larger city or a region may involve contacting all of the relevant sites and finally resorting only to the ones that actually respond. Because of the lack of census data on multilingualism due to political sensitivity, De Houwer (2004) reports on contacting principals of Dutch-medium primary schools throughout Flanders to ask them to cooperate on a large-scale survey of language use in the home. Those who were willing to participate had all of the children in the school fill out the questionnaires, totaling 18,046. Although the sample may not be representative, in that some school principals did refrain from participating for whatever reason, the study was able to reveal empirically "some idea of the incidence of multilingualism and of the languages involved" (p. 120). Deprez (1999) used a similar approach to reveal the bilingualism in the schools in Paris. Once the bilinguals are identified, follow-up work can be done.

Another type of sampling involves a *non-random* or *non-probability sampling* design. This does not follow the theory of probability in the choice of individuals from the sampling population. Such a sampling design is used when the number of people in a population is either unknown or cannot be identified. Hence it is quite suitable for bilingual research. A potential disadvantage of such a design is that it is not based on probability, and thus one cannot in principle generalize the results to the entire sampling population. However, as Milroy and Gordon (2003) point out with regard to many large-scale sociolinguistic variation studies that have employed stratified random sampling, the number of individuals in each subgroup may be so small that generalizability becomes questionable. And more often than not, such sampling ends up being judgmental, a type of non-probability sampling.

10.1002/9781444301120.ch5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In judgmental or purposive sampling, the individuals selected from a group or community are judged suitable by the researcher, usually on the basis of participation – observation of the group or community (see Codó, chapter 9 in this volume). Most studies of bilingual communities employ this method of selection. Judgmental sampling can also be used to study unique cases that are especially informative. Quota sampling may be used as an added dimension to judgmental sampling. The researcher identifies certain types of speakers, representing such social variables as age and gender, and then selects a certain quota of speakers who fit the profile.

Judgmental sampling is employed not only in studies in which the focus is on the language user but also in those that focus on the languages themselves in grammatical studies of code-switching. In her grammatical analysis of Turkish–Norwegian code-switching, Türker (2000) used data from eight bilingual speakers, three women and five men, all between the ages of 16 and 24 at the time of the recordings, who were born in Turkey and brought to Norway by their parents at various ages, and hence were of the intermediate generation judged to be representative of active code-switchers. In addition to being a judgmental sample, the sample may also contain a quota of individuals representing relevant social variables. Eze (1998) based his language contact study on data from 20 adult speakers in Nigeria fluent in Igbo and English. He states (p. 185), "The informants (9 women and 11 men) represent a variety of occupational groupings, including teachers, traders, and students, among others, and range in age from 18 to 45. All were born in Nigeria, and are highly educated, and learned Igbo as their first language and English at school age." The object was to get as representative a sample as possible. The participants in these two studies had the same native language. Proficiency in one of the languages may have an impact on the grammatical structure of the utterances exhibiting language contact. The selection of participants in both studies was guided by the theory the researchers used. Türker's theoretical framework was the Matrix Language Frame Model (Myers-Scotton, 1997, 2002) and her interest was in how Turkish as the matrix language would accommodate Norwegian as an embedded language, so she selected participants who were more proficient in Turkish. Eze's study is in the quantitative paradigm in variation theory (Poplack, 1993; Poplack & Meechan, 1998) applied to bilingual corpora (see Backus, chapter 13 in the present volume, on data banks and corpora) with an interest in loan words incorporated into Igbo speech.

Another type of non-random/non-probability sampling is referred to as *snowball sampling* (Scott, 2000). The sampling techniques discussed thus far took as a starting point the individual as a representative of a certain social variable. Snowball sampling is used as a sampling technique in examining social networks and hence takes a pre-existing group as the point of departure. This technique can be illustrated with the Filipino study (Hvenekilde & Lanza, 2001; Lanza & Svendsen, 2007) mentioned above. With the help of our field assistant, we contacted a few individuals, who were interviewed. After each interview, the participant was asked to nominate others for the study. These individuals were then interviewed and asked to nominate further participants for the study. This procedure continued, building up the selection of participants like a snowball. There are advantages and disadvantages in selecting participants through this technique. Participants are

10.1002/9781444301120.ch5 Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

easily located and as they are recommended by someone they know, access is easier for the researcher. Moreover, in studies of social networks, information gleaned from the interviews can be used to check and cross-check the structure of the participants' reported networks. On the other hand, the composition of the entire sample rests upon the choice of individuals at the first stage. If these individuals belong to a particular faction or have strong biases, the study may be biased. Moreover, it is difficult to use this technique when the sample becomes fairly large. In the Filipino study, although many participants were contacted through this procedure, the sample was supplemented by judgmental sampling in order to gain a more diversified sample.

Once we have selected the individuals from a community or group that we wish to study, we will need to consider possible factors contributing to variation in language use and form among the bilingual speakers. Will, for example, the family be the site for data collection or the school, or any other institution? Jørgensen (1998) studied bilingual Turkish–Danish children's code-switching for wielding power in a school context. He states (p. 242), "The further away from the children's everyday world, the more we find hostility towards the minority tongue, the less it is considered useful or even appropriate." Yet at the interactional level in a school setting, there were only occasional attempts to allude to the superior position of Danish (see Cashman, chapter 16 in this volume). Svennevig (2003) investigated a certain conversational structure in native/non-native interaction in consultations at public offices. Although the structure is not uncommon, the linguistic asymmetry that occurs in such an institutional setting renders the structure particularly salient. Hence, the selection of the site for data can have theoretical importance. It can reveal various aspects of the bilingual's competence and represent society in a different way than traditional correlations between linguistic and social variables such as socioeconomic status, gender, age, and so forth.

When studying a bilingual community, the researcher may wish to select a particular site for investigating language contact, acquisition, or use, merely as a point of departure. McCormick (2002) initially selected a nursery school for her study of District Six, an inner-city neighborhood in Cape Town, characterized by code-switching between English and Afrikaans. She was interested in children's code-switching ability but soon realized that she would need a more thorough grasp of language attitudes and practices in the neighborhood in order to fully understand the children's use of language. Hence, an initial sample may be extended or even modified as the research progresses and we ask further questions. Indeed, the study of one bilingual community can reveal interesting dimensions in another bilingual community, sharing, for example, the same language. Lane (1999) investigated language contact in a bilingual Finnish community in Norway while Lane (2006) investigates code-switching patterns in a bilingual Finnish community in Canada in order to understand the roles of typology and social factors in language contact patterns.

The focus in this section has been on selecting individuals in a group or community that will be representative, in order to assure the generalizability of the findings. However, we may wish to restrict our attention to a case study of a particularly interesting group, or community of practice, even though that group is not necessarily

10.1002/9781444301120.ch5. Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

representative for all similar groups. For example, language use in a particular bilingual youth community club may be an interesting and legitimate object of inquiry even though that club may differ from other bilingual youth community clubs (cf. Kallmeyer & Keim, 2003). In this case, generalizability is not an issue.

Section summary: A focus on the group or community

- *Sampling* is the process of selecting a few individuals from a larger group in order to estimate or predict aspects of the larger group.
- A sampling design may be a *random or probability* sample, or a *non-random or non-probability* sample such as judgmental or snowball. In principle, one can generalize from the former type, but not from the latter. Because of the difficulty of defining the universe from which a sample is drawn in bilingual studies, a non-random or non-probability sample is suitable for bilingual research.
- Samples can be *stratified* to reflect various social variables such as age, sex, socioeconomic status, etc.
- In *judgmental* samples, the individuals selected are judged suitable by the researcher.
- *Snowball sampling* is used in studies involving social networks.
- In a case study of a particular group, generalizability is not necessarily an issue.

5.6 Ethical Considerations

As researchers we must be aware of ethical issues involving how we treat those whom we have selected, how we collect our data, and what we do with our results. This involves all research; however, as noted above, minority groups are particularly vulnerable as our research may be used to reinforce pre-existing stereotypes, or may be misused politically or culturally (cf. Zentella, 1997; Li Wei, 2000). Svendsen (2004) reports that the parents of some of the children in her study were at first skeptical due to previous participation in a project that had negative consequences for their work situation.

Many countries will have legal safeguards for participation in research projects. These involve informed consent by the participant, often in written form (see the consent form taken from Lanza, 2004a, in the appendix at the end of this chapter). Surreptitious recording is usually illegal. The participants should clearly understand the purpose and relevance of the study. They must, furthermore, be guaranteed anonymity and must have access to the recordings made of them. Moreover, they may have the right to withdraw from a project at any time and demand that the data be destroyed.

10.1002/9781444301120.ch5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Some populations, for a number of reasons, may not feel at ease with a particular method of data collection (such as an interview), or comfortable about expressing opinions in a questionnaire. A population can in fact be tired of always being the object of study, especially if the members feel they do not get any feedback concerning the results of the study, or feel that such results may be used to their disadvantage. Sensitivity to such issues is the researcher's responsibility. Indeed the researcher's responsibility to the community has become a burning issue, particularly in sociolinguistic research. Cameron, Frazer, Harvey, Rampton, and Richardson (1992) point out that research in the social sciences is often about power relationships, and discuss the potential for developing a research process through which both the researcher and the researched can benefit. What they advocate is the "empowering approach," which gives greater power to the researched by rendering the process one of "research *on*, *for* and *with* social subjects." Every researcher of bilingualism will have to think consciously about how the researched may be involved in the research process. The ethical issues raised here should be an integral part of any work on bilingualism.

Further reading

There are many good introductory books to research methods in the social sciences, particularly aimed at graduate students and containing a chapter on sampling such as Kumar (1999) and Neuman (2003). Milroy and Gordon (2003), Chapter 2, on "Locating and Selecting Subjects," relates specifically to sociolinguistic methods. Johnstone (2000), Chapter 4, discusses "Some Legal and Ethical Issues" in sociolinguistic research. Concerning bilingualism more specifically, see Grosjean (1998) for an excellent overview of methodological and conceptual issues in studying bilinguals, including the selection of individuals, particularly from a psycholinguistic perspective. Li Wei (2000) provides a reflected discussion of methodology particularly from a sociolinguistic perspective. Chapter 3 of Lanza (2004a) presents a detailed discussion on the search for, and selection of, bilingual children and families for case studies of bilingual first language acquisition. Zentella (1997) is a comprehensive study of a bilingual community; Chapter 1 provides an interesting discussion concerning the issue of researcher identity in studying bilingualism in the community context. Fabbro (1999) and Paradis (2004) cover some methodological considerations in neurolinguistic studies of bilinguals.

10.1002/9781444301120.ch5. Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix: Consent form (from Lanza, 2004a: 349)

We agree to participate in, and to permit our child to participate in, a study of bilingual language development, conducted by Elizabeth Lanza (a Ph.D. Candidate in linguistics from Georgetown University, Washington, D.C., and NAVF fellow here in Norway) with the understanding that:

(1) The purpose of the study is to observe and describe how young children from bilingual families learn and use language under ordinary circumstances. The intent is not to change our behavior or our child's behavior;
(2) Our child will be audio-taped by E. Lanza in our home for about one hour every month for several months. After this taping session, we will record up to 90 minutes of the child interacting with family members (and/or friends). These tapes will record the child's speech and activities as she/he interacts routinely with family (and/or friends). Scheduling of taping sessions will be made at our convenience;
(3) A diary of our child's language development will be kept by us during the study. At the end of the project, the diary will be our possession; (cf. (7) below)
(4) All tapes will be listened to and analyzed only by E. Lanza, and a limited number of associates, and only for educational and scientific research purposes. The same applies to the material in the diary. At all times our identity will be kept confidential;
(5) Neither we nor any member of our family shall be identified by our actual names in any use made of the tapes or the entries in the diary;
(6) We shall have the right to listen to all audio tapes and to erase any of them or part of them;
(7) At the end of the project, E. Lanza is allowed to keep these tapes and a copy of the diary for future educational and scientific research purposes.

The collection of data of the type in this study is subject to authorization by *Datatilsynet* (cf. *Personregisterlovens* § 9). The conditions stated above in this consent form fulfill the guidelines provided by *Datatilsynet* and the Norwegian Social Science Data Service (NSD).

Signature of parents:

____

____

Signature of investigator:

____

10.1002/9781444301120.ch5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/9781444301120.ch5 by University Of Oslo, Wiley Online Library on [09/02/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License